Supplementary Materials for "Sparse Principal Component based High-Dimensional Mediation Analysis"

Yi Zhao[1], Martin A. Lindquist[1], and Brian S. Caffo[1]

[1]Department of Biostatistics, Johns Hopkins Bloomberg School of Public Health

A Causal assumptions

With single mediator, in order to estimate the causal effects from observed data, the following assumptions are imposed.

Assumption 1 The stable unit treatment value assumption (SUTVA).

Assumption 2 The structural models (1) and (2) are correctly specified.

Assumption 3 The observed outcome is one realization of the potential outcome with observed treatment assignment.

Assumption 4 The treatment X is randomized with $0 < \mathbb{P}(X = x) < 1$, i.e.,

$$\{Y(x', m), M(x)\} \perp\!\!\!\perp X.$$

Assumption 5 The mediator is ignorable, i.e.,

$$Y(x', m) \perp\!\!\!\perp M(x) \mid X = x.$$

The SUTVA definition was introduced by Rubin (1986), which contains two parts of assumptions: i) there is no multiple versions of treatment; ii) and there is "no interference" in the sense that the treatment of one subject/trial has no impact on the outcome of other subjects/trials. Assumption 2 assumes the proposed structural equation models are correctly specified, that is assuming there is no interaction between the treatment and the mediation effect (Assumption 2 in Imai et al. (2010)). Assumption 3 is also known as the "consistency" assumption (Cole and Frangakis, 2009). Assumptions 4 and 5 together is the sequential ignorability assumption in causal mediation analysis (Imai et al., 2010). These two assumptions assume there is i) no unmeasured confounding of treatment-outcome relationship, ii) no unmeasured confounding of mediator-outcome relationship, and iii) no unmeasured confounding of treatment-mediator relationship. For multiple mediators, the ignorability of mediator assumption is extended to a sequential ignorability assumption, see detailed discussion in Imai and Yamamoto (2013) for both the cases with causally independent and dependent mediators.

B Theory and Proof

B.1 Proof of Proposition 1

Proof. Without loss of generality, we assume the data is center at zero and drop the intercept term in models (5) and (6). The ordinary least square (OLS) estimator of the model coefficient in model (5) is

$$\hat{\alpha}_j^{(1)} = (\mathbf{X}^\top \mathbf{X})^{-1} \mathbf{X}^\top \tilde{\mathbf{M}}^{(1j)} = (\mathbf{X}^\top \mathbf{X})^{-1} \mathbf{X}^\top \mathbf{M} \boldsymbol{\phi}_j,$$

$$\hat{\alpha}_j^{(2)} = (\mathbf{X}^\top \mathbf{X})^{-1} \mathbf{X}^\top \tilde{\mathbf{M}}^{(2j)} = (\mathbf{X}^\top \mathbf{X})^{-1} \mathbf{X}^\top \mathbf{M}(-\boldsymbol{\phi}_j) = -(\mathbf{X}^\top \mathbf{X})^{-1} \mathbf{X}^\top \mathbf{M} \boldsymbol{\phi}_j,$$

$$\Rightarrow \quad \hat{\alpha}_j^{(1)} = -\hat{\alpha}_j^{(2)}.$$

Assume $\tau_1^2 = \mathbf{X}^\top\mathbf{X}$, $\tau_2^2 = \tilde{\mathbf{M}}^{(1j)\top}\tilde{\mathbf{M}}^{(1j)} = \boldsymbol{\phi}^\top\mathbf{M}^\top\mathbf{M}\boldsymbol{\phi}_j = \tilde{\mathbf{M}}^{(2j)\top}\tilde{\mathbf{M}}^{(2j)}$, $\rho = \mathbf{X}^\top\tilde{\mathbf{M}}^{(1j)} = \mathbf{X}^\top\mathbf{M}\boldsymbol{\phi}_j$, then $\mathbf{X}^\top\tilde{\mathbf{M}}^{(2j)} = -\rho$. The OLS estimator of β_j and γ_j in model (6) are

$$\begin{pmatrix} \hat{\gamma}_j^{(1)} \\ \hat{\beta}_j^{(1)} \end{pmatrix} = \frac{1}{\tau_1^2\tau_2^2 - \rho^2} \begin{pmatrix} \tau_2^2\mathbf{X}^\top Y - \rho\tilde{\mathbf{M}}^{(1j)\top}Y \\ \tau_1^2\tilde{\mathbf{M}}^{(1j)\top}Y - \rho\mathbf{X}^\top Y \end{pmatrix} = \frac{1}{\tau_1^2\tau_2^2 - \rho^2} \begin{pmatrix} \tau_2^2\mathbf{X}^\top Y - \mathbf{X}^\top\mathbf{M}\boldsymbol{\phi}_j\boldsymbol{\phi}_j^\top\mathbf{M}^\top Y \\ \tau_1^2\boldsymbol{\phi}_j^\top\mathbf{M}^\top Y - \boldsymbol{\phi}_j^\top\mathbf{M}^\top\mathbf{X}\mathbf{X}^\top Y \end{pmatrix},$$

and

$$\begin{pmatrix} \hat{\gamma}_j^{(2)} \\ \hat{\beta}_j^{(2)} \end{pmatrix} = \frac{1}{\tau_1^2\tau_2^2 - \rho^2} \begin{pmatrix} \tau_2^2\mathbf{X}^\top Y + \rho\tilde{\mathbf{M}}^{(2j)\top}Y \\ \tau_1^2\tilde{\mathbf{M}}^{(2j)\top}Y + \rho\mathbf{X}^\top Y \end{pmatrix} = \frac{1}{\tau_1^2\tau_2^2 - \rho^2} \begin{pmatrix} \tau_2^2\mathbf{X}^\top Y - \mathbf{X}^\top\mathbf{M}\boldsymbol{\phi}_j\boldsymbol{\phi}_j^\top\mathbf{M}^\top Y \\ -\tau_1^2\boldsymbol{\phi}_j^\top\mathbf{M}^\top Y + \boldsymbol{\phi}_j^\top\mathbf{M}^\top\mathbf{X}\mathbf{X}^\top Y \end{pmatrix}.$$

$$\Rightarrow \quad \hat{\beta}^{(1j)} = -\hat{\beta}^{(2j)}, \quad \hat{\gamma}^{(1j)} = \hat{\gamma}^{(2j)}.$$

Therefore, the estimate of the indirect effect

$$\widehat{\mathrm{IE}}^{(1j)} = \hat{\alpha}^{(1j)}\hat{\beta}^{(1j)} = \hat{\alpha}^{(2j)}\hat{\beta}^{(2j)} = \widehat{\mathrm{IE}}^{(2j)}.$$

The estimate of the direct and indirect effects are sign invariant with respect to the loadings. $\square$

B.2 A Sketch of Asymptotic Properties of the Estimators in Algorithm 1

We propose a sparse principal component mediation analysis approach to enable intuitive interpretations. The sparsification and dependence removal procedures modify the mediator PCs, and thus change the estimate of the causal parameters. In this section, we provide a sketch of asymptotic convergence of the estimators obtained in Algorithm 1 and the estimators from the original mediator PCs.

As the sparsified loadings are not orthogonal, in the algorithm, we include a linear dependence removal procedure through the following regression model: for $k = 1, \ldots, q$,

$$\check{M}_i^{(k)} = \pi_0 + \pi_1 X_i + \check{\mathbf{M}}_i^{(1,\ldots,k-1)\top}\boldsymbol{\Pi}_{1,\ldots,k-1} + \tau_i^{(k)} = \mathbf{X}_i^\top\boldsymbol{\pi}_k + \check{\mathbf{M}}_i^{(1,\ldots,k-1)\top}\boldsymbol{\Pi}_{1,\ldots,k-1} + \tau_i^{(k)}.$$

Put it under the matrix form, and we have the solution for $\boldsymbol{\Pi}_{1,\ldots,k-1}$ as

$$\hat{\boldsymbol{\Pi}}_{1,\ldots,k-1} = \left[\check{\mathbf{M}}^{(1,\ldots,k-1)\top}(\mathbf{I} - \mathbf{P}_\mathbf{X})\check{\mathbf{M}}^{(1,\ldots,k-1)}\right]^{-1}\check{\mathbf{M}}^{(1,\ldots,k-1)\top}(\mathbf{I} - \mathbf{P}_\mathbf{X})\check{\mathbf{M}}^{(k)},$$

where $\mathbf{P}_\mathbf{X} = \mathbf{X}(\mathbf{X}^\top\mathbf{X})^{-1}\mathbf{X}^\top$ is the projection matrix of $\mathbf{X}$. The new mediator is

$$\check{\mathbf{M}}^{(k\cdot 1,\dots,k-1)} = \check{\mathbf{M}}^{(k)} - \check{\mathbf{M}}^{(1,\dots,k-1)}\hat{\mathbf{\Pi}}_{1,\dots,k-1} = (\mathbf{I} - \mathbf{Q}_{\check{\mathbf{M}}^{(1,\dots,k-1)}})\check{\mathbf{M}}^{(k)},$$

where

$$\mathbf{Q}_{\check{\mathbf{M}}^{(1,\dots,k-1)}} = \check{\mathbf{M}}^{(1,\dots,k-1)}\left[\check{\mathbf{M}}^{(1,\dots,k-1)\top}(\mathbf{I}-\mathbf{P}_\mathbf{X})\check{\mathbf{M}}^{(1,\dots,k-1)}\right]^{-1}\check{\mathbf{M}}^{(1,\dots,k-1)\top}(\mathbf{I}-\mathbf{P}_\mathbf{X}).$$

Let $\hat{\mathbf{a}}_k$ and $\hat{\mathbf{a}}_{k\cdot 1,\dots,k-1}$ denote the estimated coefficient in model (5) with $\check{\mathbf{M}}^{(k)}$ and $\check{\mathbf{M}}^{(k\cdot 1,\dots,k-1)}$ as the dependent variable, respectively. We first consider the case when $k=2$. Then

$$\hat{\mathbf{a}}_2 = (\mathbf{X}^\top\mathbf{X})^{-1}\mathbf{X}^\top\check{\mathbf{M}}^{(2)},$$

$$\hat{\mathbf{a}}_{2\cdot 1} = (\mathbf{X}^\top\mathbf{X})^{-1}\mathbf{X}^\top\check{\mathbf{M}}^{(2\cdot 1)} = (\mathbf{X}^\top\mathbf{X})^{-1}\mathbf{X}^\top(\mathbf{I}-\mathbf{Q}_{\check{\mathbf{M}}^{(1)}})\check{\mathbf{M}}^{(2)},$$

$\Rightarrow$

$$\begin{aligned}
\hat{\mathbf{a}}_2 - \hat{\mathbf{a}}_{2\cdot 1} &= (\mathbf{X}^\top\mathbf{X})^{-1}\mathbf{X}^\top\check{\mathbf{M}}^{(2)} - (\mathbf{X}^\top\mathbf{X})^{-1}\mathbf{X}^\top(\mathbf{I}-\mathbf{Q}_{\check{\mathbf{M}}^{(1)}})\check{\mathbf{M}}^{(2)} \\
&= (\mathbf{X}^\top\mathbf{X})^{-1}\mathbf{X}^\top\mathbf{Q}_{\check{\mathbf{M}}^{(1)}}\check{\mathbf{M}}^{(2)} \\
&= (\mathbf{X}^\top\mathbf{X})^{-1}\mathbf{X}^\top\check{\mathbf{M}}^{(1)}\left[\check{\mathbf{M}}^{(1)\top}(\mathbf{I}-\mathbf{P}_\mathbf{X})\check{\mathbf{M}}^{(1)}\right]^{-1}\check{\mathbf{M}}^{(1)\top}(\mathbf{I}-\mathbf{P}_\mathbf{X})\check{\mathbf{M}}^{(2)}.
\end{aligned}$$

After sparsifying the loadings, $\check{\mathbf{M}}^{(1)}$ and $\check{\mathbf{M}}^{(2)}$ are not conditionally independent given X. But for $k \neq j$, the original PC mediators $\tilde{\mathbf{M}}^{(k)} = \mathbf{M}\hat{\mathbf{v}}_k$ and $\tilde{\mathbf{M}}^{(j)} = \mathbf{M}\hat{\mathbf{v}}_j$ are. Therefore,

$$\tilde{\mathbf{M}}^{(1)\top}(\mathbf{I}-\mathbf{P}_\mathbf{X})\tilde{\mathbf{M}}^{(2)} = \left[(\mathbf{I}-\mathbf{P}_\mathbf{X})\tilde{\mathbf{M}}^{(1)}\right]^\top\left[(\mathbf{I}-\mathbf{P}_\mathbf{X})\tilde{\mathbf{M}}^{(2)}\right] = \mathcal{O}_p\left(\sqrt{\frac{1}{n}}\right).$$

We sparsify the loadings through an ℓ_1-type regularization. Considering lasso penalty, Chatterjee (2013) showed that when the design matrix and the true coefficient parameters are bounded, the expected mean squared prediction error converges to zero at the rate of $\sqrt{\log p/n}$. Under our problem setting, this is saying

$$\mathbb{E}\left(\widehat{\mathrm{MSPE}}(\hat{\mathbf{w}}_k)\right) = \mathcal{O}_p\left(\sqrt{\frac{\log(p)}{n}}\right),$$

where $\widehat{\mathrm{MSPE}}(\hat{\mathbf{w}}_k) = \|\mathbf{E}\hat{\mathbf{v}}_k - \mathbf{E}\hat{\mathbf{w}}_k\|_2^2$. Assuming $\sum_{i=1}^n X_i^2/n \to \text{const} < \infty$,

$$\mathbb{E}\left(\|\mathbf{M}\hat{\mathbf{v}}_k - \mathbf{M}\hat{\mathbf{w}}_k\|_2^2\right) = \mathcal{O}_p\left(\sqrt{\frac{\log(p)}{n}}\right).$$

Therefore,

$$\mathbb{E}(\|\hat{\mathbf{a}}_2 - \hat{\mathbf{a}}_{2\cdot 1}\|_2^2) = \mathcal{O}_p\left(\sqrt{\frac{\log(p)}{n}}\right).$$

Let $\hat{\boldsymbol{\alpha}}_k$ denote the estimated coefficient in model (5) with $\tilde{\mathbf{M}}^{(k)}$ as the dependent variable.

$$\begin{aligned}
\mathbb{E}(\|\hat{\boldsymbol{\alpha}}_2 - \hat{\mathbf{a}}_2\|_2^2) &= \mathbb{E}(\|(\mathbf{X}^\top\mathbf{X})^{-1}\mathbf{X}^\top\tilde{\mathbf{M}}^{(2)} - (\mathbf{X}^\top\mathbf{X})^{-1}\mathbf{X}^\top\check{\mathbf{M}}^{(2)}\|_2^2) \\
&= \mathbb{E}(\|(\mathbf{X}^\top\mathbf{X})^{-1}\mathbf{X}^\top\mathbf{M}\hat{\mathbf{v}}_2 - (\mathbf{X}^\top\mathbf{X})^{-1}\mathbf{X}^\top\mathbf{M}\hat{\mathbf{w}}_2\|_2^2) \\
&= \mathcal{O}_p\left(\sqrt{\frac{\log(p)}{n}}\right).
\end{aligned}$$

Therefore,

$$\mathbb{E}(\|\hat{\mathbf{a}}_{2\cdot 1} - \hat{\boldsymbol{\alpha}}_2\|_2^2) \leq \mathbb{E}(\|\hat{\mathbf{a}}_{2\cdot 1} - \hat{\mathbf{a}}_2\|) + \mathbb{E}(\|\hat{\mathbf{a}}_2 - \hat{\boldsymbol{\alpha}}_2\|) = \mathcal{O}_p\left(\sqrt{\frac{\log(p)}{n}}\right).$$

For model (6),

$$\begin{aligned}
\hat{\beta}_2 &= [\tilde{\mathbf{M}}^{(2)\top}(\mathbf{I} - \mathbf{P}_{\mathbf{X}})\tilde{\mathbf{M}}^{(2)}]^{-1}\tilde{\mathbf{M}}^{(2)\top}(\mathbf{I} - \mathbf{P}_{\mathbf{X}})\mathbf{Y}, \\
\hat{b}_2 &= [\check{\mathbf{M}}^{(2)\top}(\mathbf{I} - \mathbf{P}_{\mathbf{X}})\check{\mathbf{M}}^{(2)}]^{-1}\check{\mathbf{M}}^{(2)\top}(\mathbf{I} - \mathbf{P}_{\mathbf{X}})\mathbf{Y}, \\
\hat{b}_{2\cdot 1} &= [\check{\mathbf{M}}^{(2\cdot 1)\top}(\mathbf{I} - \mathbf{P}_{\mathbf{X}})\check{\mathbf{M}}^{(2\cdot 1)}]^{-1}\check{\mathbf{M}}^{(2\cdot 1)\top}(\mathbf{I} - \mathbf{P}_{\mathbf{X}})\mathbf{Y} \\
&= [\check{\mathbf{M}}^{(2)\top}(\mathbf{I} - \mathbf{Q}_{\check{\mathbf{M}}^{(1)}}^\top)(\mathbf{I} - \mathbf{P}_{\mathbf{X}})(\mathbf{I} - \mathbf{Q}_{\check{\mathbf{M}}^{(1)}})\check{\mathbf{M}}^{(2)}]^{-1}\check{\mathbf{M}}^{(2)\top}(\mathbf{I} - \mathbf{Q}_{\check{\mathbf{M}}^{(1)}}^\top)(\mathbf{I} - \mathbf{P}_{\mathbf{X}})\mathbf{Y}.
\end{aligned}$$

First, using the results from above,

$$\mathbb{E}(\|\hat{\beta}_2 - \hat{b}_2\|_2^2) = \mathcal{O}_p\left(\sqrt{\frac{\log(p)}{n}}\right).$$

Taking a similar strategy, we can also have that

$$\mathbb{E}(\|\hat{b}_2 - \hat{b}_{2\cdot 1}\|_2^2) = \mathcal{O}_p\left(\sqrt{\frac{\log(p)}{n}}\right),$$

and thus

$$\mathbb{E}(\|\hat{\beta}_2 - \hat{b}_{2\cdot 1}\|_2^2) = \mathcal{O}_p\left(\sqrt{\frac{\log(p)}{n}}\right).$$

Therefore, for the indirect effect

$$\mathbb{E}(\|\hat{\alpha}_2\hat{\beta}_2 - \hat{a}_{2\cdot 1}\hat{b}_{2\cdot 1}\|_2^2) = \mathcal{O}_p\left(\sqrt{\frac{\log(p)}{n}}\right),$$

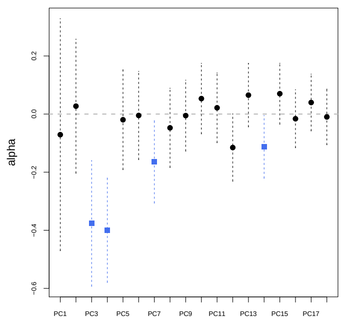
where $\hat{\alpha}_2$ and $\hat{a}_{2\cdot 1}$ are the second element of $\hat{\boldsymbol{\alpha}}_2$ and $\hat{\mathbf{a}}_{2\cdot 1}$, respectively. For higher order PCs, the same conclusion follows. The convergence of these estimators are examined through simulation study presented in Section D.

C Additional fMRI Study Results

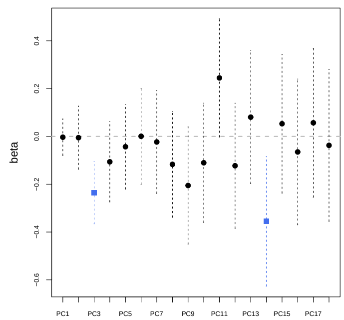
Figure C.1 shows the estimate of model coefficient and the indirect effect (IE) of the first 18 PCs in both the PCA and sparse PCA (SPCA) based mediation analysis. From the figure, only PC3 yields significant positive mediation effect with significant negative α and β estimates.

Figure C.2 shows the loadings of PC3. From the figure, we do not observe any clear patterns within each functional module. If following a thresholding step, several regions in the visual cortex would remain with a positive loading. The estimate of α for PC3 is negative indicating that these visual regions are deactivated in the classification task compared to the baseline. Further, the activation of these visual areas will shorten the reaction time as the estimate β is negative. These conclusions depart from existing findings in visual cortex when related to sensory processing and motor response (Poldrack et al., 2001; Aron et al., 2004; Yarkoni et al., 2009).

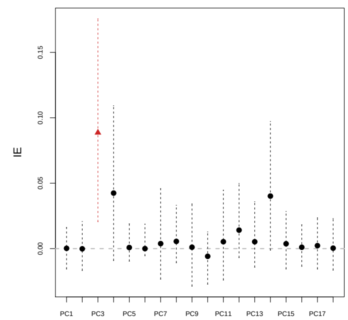
Figure C.3 presents the river plot of the sparse loadings of PC3. The downstream is separated for positive and negative loadings. From the figure, the visual regions (lateral, medial and occipital pole) yield negative loadings; and the positive loadings are contributed mainly by auditory, default mode network (DMN), executive control (EC), and frontoparietal cortex. The whole cerebellum



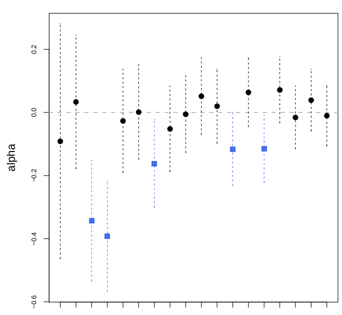
(a) α estimate in PCA analysis.



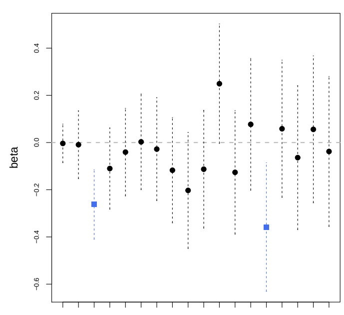
(b) β estimate in PCA analysis.



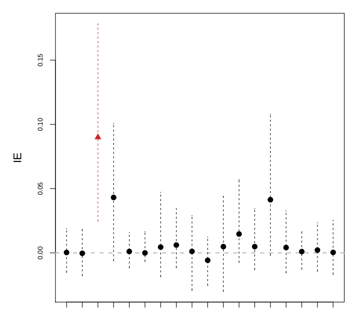
(c) IE estimate in PCA analysis.



(d) α estimate in SPCA analysis.



(e) β estimate in SPCA analysis.



(f) IE estimate in SPCA analysis.

Figure C.1: Estimate of model coefficients α and β, and the indirect effect (IE) in the (a)&(b)&(c) PCA based and (d)&(e)&(f) sparse PCA (SPCA) based analysis of the first 18 PCs.

module is penalized to zero.

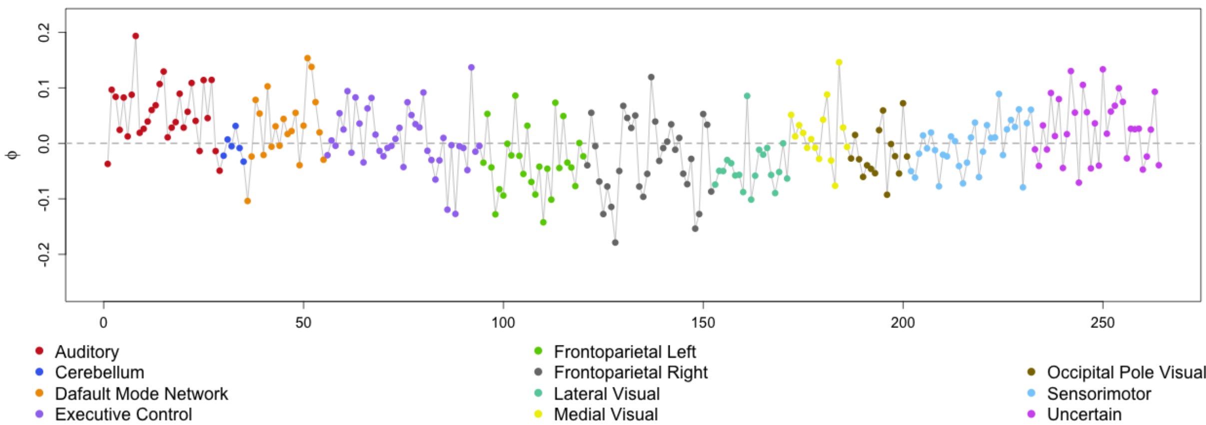


Figure C.2: The loadings of PC3. The x-axis shows the index of the 264 brain regions which are colored by functional modules.

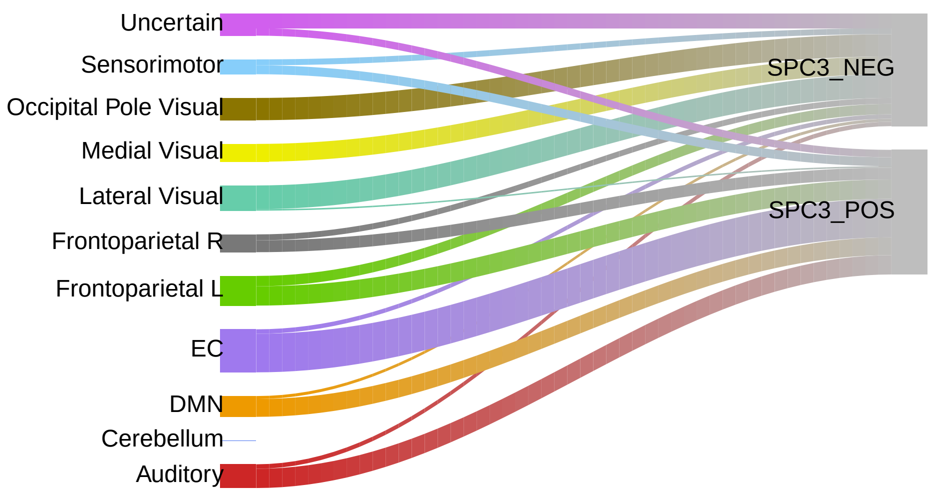


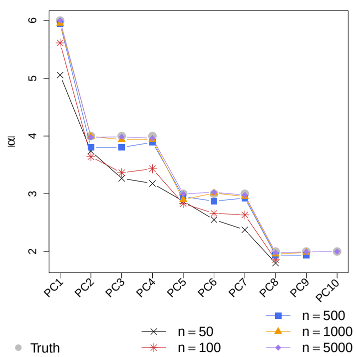
Figure C.3: River plot of the sparse approximation of PC3. The downstream is divided into positive (POS) and negative (NEG) loadings. DMN: default mode network, EC: executive control, L: left, R: right.

D Simulation Study

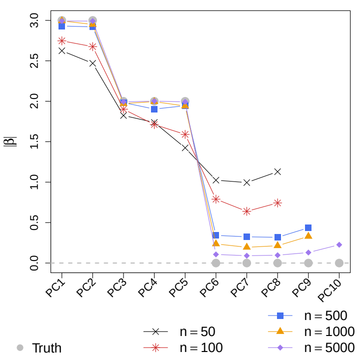
The simulated data is generated following models (5) and (6), and then the mediators are linearly transformed back to the original space. In the simulation, $p = 200$. The orthogonal matrix $\mathbf{\Phi}$ is generated with sparse loadings, and the eigenvalues in $\mathbf{\Lambda}$ decay exponentially. The samples sizes are set to be $n = 50$, 100, 500, 1000, 5000 to contain both cases where $n > p$ and $n < p$. The simulation is repeated 200 times. We compare the performance of (1) the PCA based mediation analysis (PCA) and (2) the proposed sparse PC based mediation analysis (SPCA). The number of PCs is chosen so that at least 75% of the data variation is explained. The sparse PC based mediation analysis is performed following Algorithm 1.

Figure D.1 shows the estimate of model coefficients, as well as the indirect effect (IE) under different values of n with $p = 200$. Since the estimates of α and β are sign nonidentifiable, we compare the estimate of their absolute values. From the figure it is clear that, as the number of observations increases, the estimate from both methods converge to the truth. The PCA approach achieves lower bias in estimating $|\alpha|$ and $|\beta|$, while the difference between the two methods diminishes when estimating the indirect effect. The same conclusion can be drawn for the case with $p = 20$ and $p = 500$ and sample size $n = 50$ (Figures D.3a and D.3d).

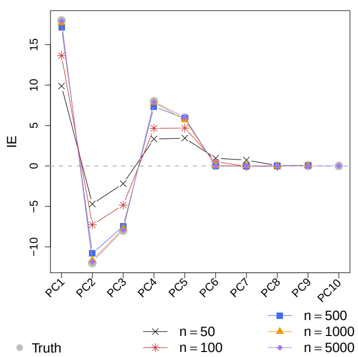
Figure D.2, as well as Figures D.3b–D.3c and D.3e–D.3f, presents the coverage probability (CP) and the power of the indirect effect. The confidence intervals are obtained from 500 bootstrap samples at the significance level of 0.95 following the procedure proposed in Section 3.4. From the figures, the CP of those PCs with nonzero IE is slightly off, while the power increases to one as n increases; for those PCs with zero IE, both the CP and the Type I error are well controlled at the designated level. When comparing the PCA and SPCA approaches, no significant difference is observed.



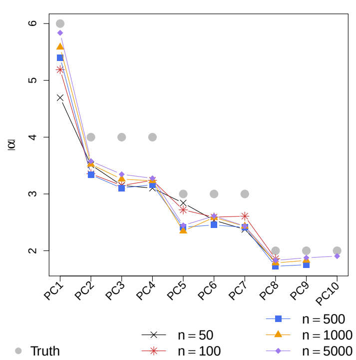
(a) Estimate of $|\alpha|$ under PCA.



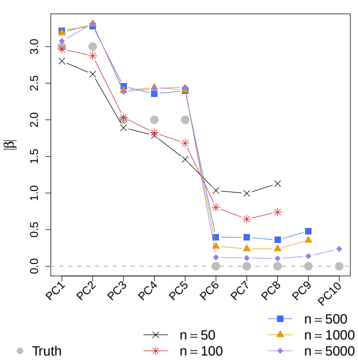
(b) Estimate of $|\beta|$ under PCA.



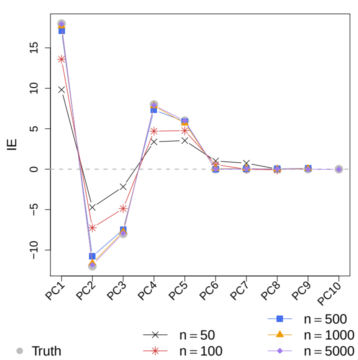
(c) Estimate of IE under PCA.



(d) Estimate of $|\alpha|$ under SPCA.

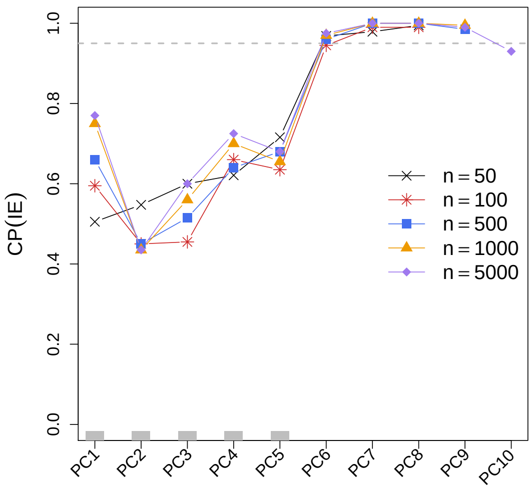


(e) Estimate of $|\beta|$ under SPCA.

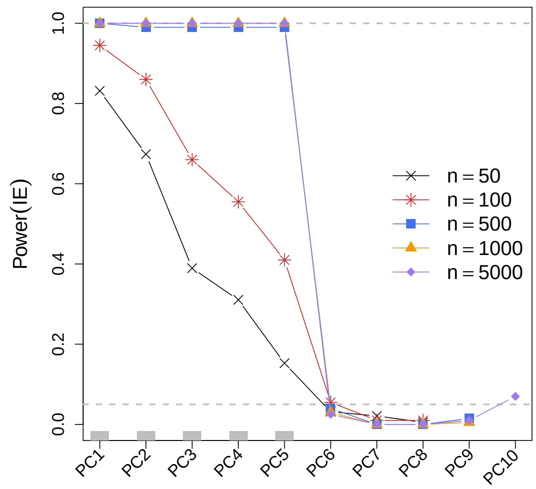


(f) Estimate of IE under SPCA.

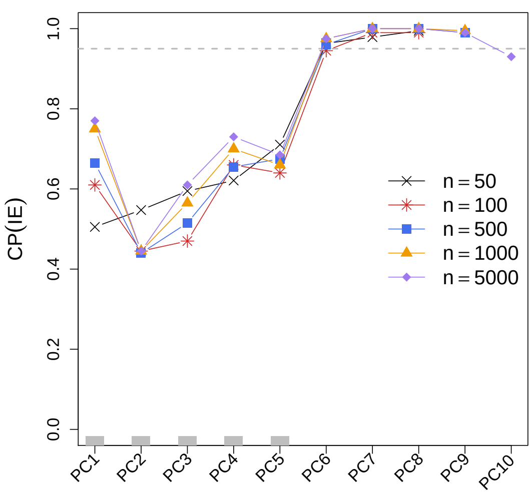
Figure D.1: Estimate of $|\alpha|$, $|\beta|$ and the indirect effect (IE) over 200 replications under different numbers of observations with $p = 200$. The gray dots are the true parameter values.



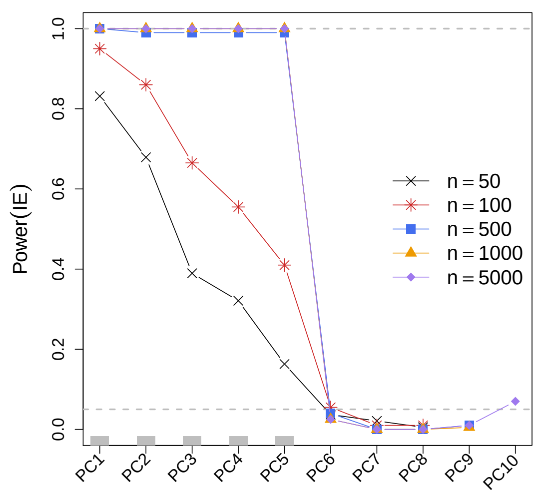
(a) CP of IE under PCA.



(b) Power of IE under PCA.

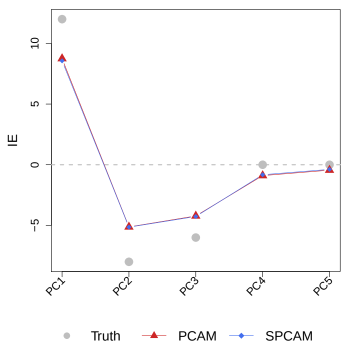


(c) CP of IE under SPCA.

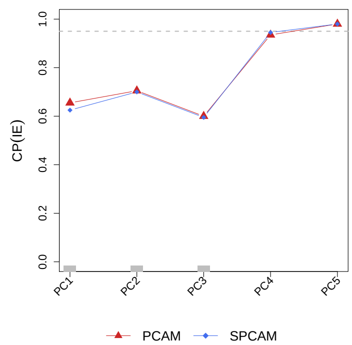


(d) Power of IE under SPCA.

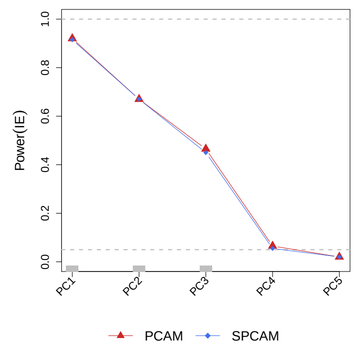
Figure D.2: Coverage probability (CP) and power of the indirect effect (IE) over 200 replications under different numbers of observations with $p = 200$, where the confidence interval is obtained from 500 bootstrap samples at the significance level of 0.95. The gray dots indicate a PC with nonzero IE effect.



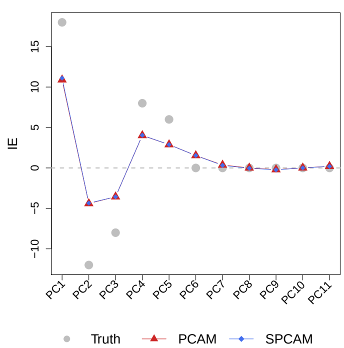
(a) Estimate of IE with $p = 20$.



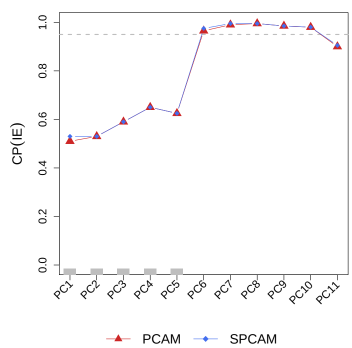
(b) CP of IE with $p = 20$.



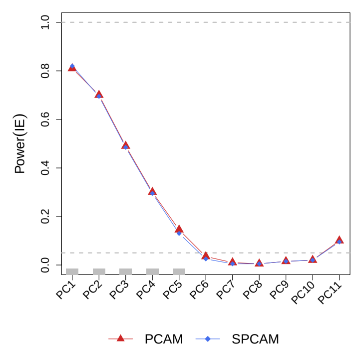
(c) Power of IE with $p = 20$.



(d) Estimate of IE with $p = 500$.



(e) CP of IE with $p = 500$.



(f) Power of IE with $p = 500$.

Figure D.3: Estimate, coverage probability (CP) and power of the indirect effect (IE) over 200 replications with $n = 50$ and (a)–(c) $p = 20$ and (d)–(f) $p = 500$. The gray dots in (a) and (d) are the true parameter values; and the gray dots in (b)–(c) and (e)–(f) indicate a PC with nonzero IE effect.

References

Aron, A. R., Shohamy, D., Clark, J., Myers, C., Gluck, M. A., and Poldrack, R. A. (2004). Human midbrain sensitivity to cognitive feedback and uncertainty during classification learning. *Journal of neurophysiology*, 92(2):1144–1152.

Chatterjee, S. (2013). Assumptionless consistency of the lasso. *arXiv preprint arXiv:1303.5817*.

Cole, S. R. and Frangakis, C. E. (2009). The consistency statement in causal inference: a definition or an assumption? *Epidemiology*, 20(1):3–5.

Imai, K., Keele, L., and Yamamoto, T. (2010). Identification, inference and sensitivity analysis for causal mediation effects. *Statistical Science*, pages 51–71.

Imai, K. and Yamamoto, T. (2013). Identification and sensitivity analysis for multiple causal mechanisms: Revisiting evidence from framing experiments. *Political Analysis*, 21(2):141–171.

Poldrack, R. A., Clark, J., Pare-Blagoev, E., Shohamy, D., Moyano, J. C., Myers, C., and Gluck, M. A. (2001). Interactive memory systems in the human brain. *Nature*, 414(6863):546–550.

Rubin, D. (1986). Statistics and causal inference: Comment: Which ifs have causal answers. *Journal of the American Statistical Association*, 81:961.

Yarkoni, T., Barch, D. M., Gray, J. R., Conturo, T. E., and Braver, T. S. (2009). Bold correlates of trial-by-trial reaction time variability in gray and white matter: a multi-study fmri analysis. *PLoS One*, 4(1):e4257.